The Communicative Model of Disinformation : A Literature Note

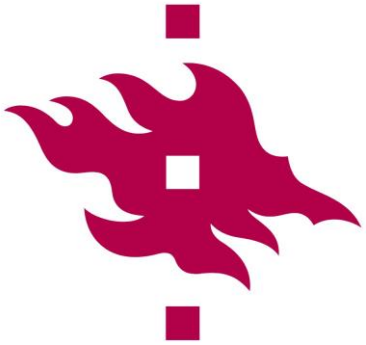
Hillebrandt, Maarten Zbigniew

University of Helsinki, Faculty of Law

2021-02

Hillebrandt , M Z 2021 ' The Communicative Model of Disinformation : A Literature Note ' Legal Studies Research Paper Series , no. 65 , University of Helsinki, Faculty of Law , Helsinki , pp. 1-15 .

http://hdl.handle.net/10138/332810



UNIVERSITY OF HELSINKI
FACULTY OF LAW

The Communicative Model of Disinformation: A Literature Note

MAARTEN HILLEBRANDT

LEGAL STUDIES RESEARCH PAPER SERIES
Paper No 65

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI
OIKEUSTIETEELLINEN TIEDEKUNTA
JURIDISKA FAKULTETEN
FACULTY OF LAW

The Communicative Model of Disinformation: A Literature Note

Maarten Hillebrandt, University of Helsinki[1]
maarten.hillebrandt@helsinki.fi

February, 2021

Abstract
In recent years, academic research and policy circles alike frequently identify disinformation and fake news as a growing problem in western democracies. This has prompted calls for regulatory intervention. In the name of protecting the circulation of factually correct information and truth, and to protect and facilitate public debate, many public authorities are proposing steps for the regulation of information flows or their platforms. Before the appropriateness of regulatory measures however can be properly assessed, a more fine-grained understanding of the phenomenon of disinformation is required. In this light, this note discusses some recent academic literature, in search of answers to three topical questions from the perspective of policy makers: (1) Does the online mode of communication alter the nature and functioning of disinformation? (2) How do the institutions for creating (and maintaining trust in) public information relate to disinformation? and (3) How do motives other than malignant intentions cause or exacerbate the disinformation phenomenon? The note relies on the concepts of 'information ecologies' (Nardi and O'Day, 1999) and 'flat ontologies' (Latour, 2005) as heuristic devices to structure recent academic insights regarding disinformation. Accordingly, disinformation is approached as a communicative phenomenon consisting of an 'assemblage' of people, practices, values, and technologies. The note describes the basic features of the late modern disinformation phenomenon, discussing in turn the actors, technological features, and drivers that are implicated in it.

[1] This working paper was written in the context of the research project "Transparency in the EU: From Reaction Manifesto?" (TrUE), funded by Academy of Finland.

1. Introduction

In recent years, disinformation and fake news are frequently identified as a growing problem in western democracies. A quick Google Scholar search suggests that the disinformation phenomenon is receiving a growing amount of academic attention (see table). This has prompted calls for regulatory intervention. In the name of protecting the circulation of factually correct information and truth, and to protect and facilitate public debate, many public authorities are proposing steps for the regulation of information flows or their platforms. Social media platforms have begun to take steps to address complaints that they lend themselves for the widespread dissemination of false information (Guardian, 2018; NBC, 2020). Some of these measures have been criticised by public authorities for being ad hoc and arbitrary (Reuters, 2020; Bloomberg, 2021), while some authorities have even begun to take steps to curb such self-regulatory interventions (BBC, 2021). Overall, calls for regulatory intervention by public authorities are sounding evermore loudly, with some western democracies, including Germany and Spain, having already pressed ahead (Alemanno, 2018; Tucker et al. 2018: 62).

Proposed measures differ in their approach. While some policy makers advocate stepping up public awareness campaigns and introducing further contextualisation measures (such as surrounding questionable information by alternative news sources), others propose more heavy-handed intervention –including various positive (e.g. due diligence and disclosure) and negative (prohibitions on publishing illegal or removing legal materials) obligations for social media platforms–, enforcement –including fines or temporary shutdowns–, and structural measures –including the formation of disinformation units identifying commonly circulating disinformation, and discussions on the further curbing of social media giants' market activities. In the context of the EU, proponents of open government have also drawn a connection between disinformation and efforts to make decision making more transparent.[2]

As can be deduced from the above-cited proposals, the purport of some of these measures can be highly intervening, touching upon various fundamental principles of the liberal democratic constitutional order, including the freedom of speech and information, free enterprise, and the taboo on direct government intervention into the activities of private citizens and media organisations (e.g. Stark and Stegmann, 2020; McGonagle, 2017).

Year	*Number*
2000	924
2005	1710
2010	2430
2015	3290
2020	7900

Table: Number of hits under the search term 'disinformation' on Google Scholar.

Before the appropriateness of regulatory measures however can be properly assessed, a more fine-grained understanding of the phenomenon of disinformation is required. Various questions require an answer from a legal point of view. What is disinformation, and to what extent can it be objectively defined? Which actors

[2] See, e.g., the comments made by Finnish Ambassador to the EU Marja Rislakki at the Seminar on the Future of EU Transparency, held on 24 September 2019 and co-organised by the Finnish EU Presidency and the TrUE project (see footnote 1 above), https://eu2019.fi/en/events/2019-09-24/seminar-on-the-future-of-eu-transparency.

are involved in its creation and dissemination, and what roles to they play in the process? But also questions concerning its material preconditions, functioning, and spread are important, if any regulation that is introduced is to have the desired effect. Such questions require insights from beyond the field of (information) law, to include communication sciences, sociology, and media and technology studies.

This note discusses some of the recent literature from these fields on the disinformation phenomenon. It focusses on three questions that are of interest from the perspective of policy makers: (1) Does the online mode of communication alter the nature and functioning of disinformation? (2) How do the institutions for creating (and maintaining trust in) public information relate to disinformation? and (3) How do motives other than malignant intentions cause or exacerbate the disinformation phenomenon? It does so by first discussing the concept of disinformation, and second, describing the central characteristics of the disinformation ecosystem, consisting of actors, technological features, and drivers.

2. Conceptual preliminaries

The phenomenon that we describe in this note as 'disinformation' has been named differently in the academic literature, to connote a variety of phenomena. Tucker and colleagues (2018), in a review of the academic literature, define disinformation in a broad manner as "the types of information that one could encounter online that could possibly lead to misperceptions about the actual state of the world" (Tucker et al., 2018: 3). As they point out, there is no consensus about what phenomena come under this definition, with candidates including disinformation and misinformation, online propaganda, hyperpartisan information, fake news, clickbait, rumours, and conspiracy theories (ibid: 55). The richness in terminology reveals a plethora of empirical phenomena, and suggests a degree of interrelation, without however explaining the nature of this interrelation.

Of the various terms mentioned by Tucker and colleagues, probably the most frequently invoked in the academic literature are *fake news* and *disinformation*. A straight-forward definition of fake news is provided by Fletcher and colleagues, who speak of "…sites that independent fact-checkers have shown to publish demonstrably false news and information, whether for profit or for ideological/political purposes" (Fletcher et al., 2018: 1). Disinformation in turn has been described as "intentional falsehoods spread as news stories or simulated documentary formats to advance political goals" (Bennett and Livingston, 2018: 124, see also Martens et al., 2018). These definitions of fake news and disinformation, which appear largely interchangeable, include some problematic presuppositions:

- The locus of fake news is intuitively and quickly associated with online communication. However, claims of fake news are not necessarily limited to the online sphere, but can also refer, inter alia, to print and television communication (European Commission, 2017 in Alemanno, 2018: 2).
- It is assumed that neutral criteria can be established to confirm the falsehood of fake news reports. As Corner however points out, the label of 'fake news' is easily attached to reports that conflict with official government pronouncements (Corner, 2017: 1100). The question of definition is thus riddled with power relations, and forms part of a wider process of information and knowledge validation (Conrad, 2020; Kofman, 2018).

- The definitions presuppose that falsehood is brought about deliberately, to deceive and manipulate (MacGonagle, 2017: 203; Conrad, 2020: 2; Allcott and Gentzkow, 2016, in Alemanno, 2018: 2). However, the possibility cannot be discounted that the spread of disinformation results from a complex interaction of factors that includes part calculated, part indirect drivers (Dahlgren, 2018).

Still, these shortcomings may also serve as the starting point for important questions surrounding talk about disinformation.

First, the strong focus on online activity in much disinformation research reflects an apparent acceleration in such activity. In particularly, the advent of large social media companies acting as algorithm-driven information platforms has been problematised for lending disinformation the high visibility and virality that it lacked hitherto. Yet the enduring presence of disinformation in the offline sphere also throws up the question whether this phenomenon is really as distinctive from previous forms of misleading public information as some make it out to be (MacGonagle, 2017: 205-7).

Second, the seemingly simple task of identifying disinformation belies the relative complexity of information concerning public matters. The term 'post-truth' is sometimes invoked to indicate the dwindling relevance of 'objective reality'. Critics have pointed towards 'postmodern' ideas of reality as a culprit in the casualisation of references to truth and objectivity (Corner, 2017: 1100, 1104; Conrad, 2020: 5). However, a distinction must be made between facts and truth, which each give rise to different epistemological conflicts. While the term 'fact' refers to a commonly accepted observation of reality, 'truth' refers to intersubjectively accepted understandings of reality, grounded in facts (Conrad, 2020: 3). As a metaphor, one can think of a connect-the-dots puzzle, in which the established points (facts) are connected by lines in a way that unveils an image that makes sense to the observer (truth). Postmodern critique may have been misconstrued as arguing that reality does not and cannot exist, whereas in fact it has been mainly concerned with the processes by which citizens and society at large seek to establish and validate representations of reality. It thus approaches the identification of both facts and truths as ultimately a socially negotiated process (Kofman, 2018). This shifts the focus towards the institutions engaged in information production, and the manner in which they gain or lose the public's trust. These processes may involve aspects that are oriented on affective, rather than rationalistic criteria of acceptance (Dahlgren, 2018: 25; also Michailidou et al., 2020: 17).

Third, the criterion of deliberateness places the focus on the malignant intentions of those spreading disinformation, whereby the intention is to deceive in order to confuse and manipulate. As suggested above, in reality, disinformation is likely to proliferate due to the actions of a variety of actors with different interests at stake. This goes beyond state-malignant interests to include financial motives, ideological drive, systemic mistrust, entertainment, and psychological factors associated with information consumption styles (Vasu et al.,2018: 5; Michailidou et al., 2018).

In sum, conceptualisations of the disinformation phenomenon open up three areas for further inquiry: (1) whether online communication alters its nature and functioning, and if so, how; (2) how the institutions for creating (and maintaining trust in) public information relate to disinformation; and (3) how motives other than malignant intentions may cause or exacerbate the disinformation phenomenon. The answer to these

questions will have an impact on the selected approach of potential future regulatory instruments, with regard to their addressee, purpose and manner of intervention, and finally, choice of legal instruments (e.g. criminal legal, market regulatory, or soft 'nudging'). The next section of this note explores these questions in more detail. It considers, in turn, the actors, the infrastructure, and the likely drivers and motives associated with disinformation.

3. Mapping the disinformation ecosystem

This note uses the metaphor of an 'ecosystem' as a heuristic device for describing the communicative constellation within which late modern disinformation exists. Our use of this concept lies close to the metaphor of 'information ecology' coined by the anthropologists Nardi and O'Day (1999). Their notion refers to the whole of "people, practices, values, and technologies in a particular local environment" engaged in information technology-driven activities. As Nardi and O'Day argue, the use of the ecology metaphor allows the analyst to highlight the systemic, diverse, and evolutionary character of information use, as well its reliance on what they call 'keystone species' and local characteristics, knowledge, and values (Nardi and O'Day, 1999). However, given the long chains of communication that are entailed in late modern communication media (including and perhaps especially social media), describing the phenomenon as 'local' seems a stretch. Latour (2005) has sought to bridge the divide between the local and the universal by referring to 'flat ontologies', i.e., networks of shared perceptions of reality that are held together by 'assemblages' of facts, artefacts, and relations. In this note, we speak of (dis)information ecosystems (rather than ecologies), in a conceptualisation that combines elements of both approaches. We highlight those factors that guarantee the viability of disinformation ecosystems which potentially span large physical and interpersonal distances. We divide these factors up into actors, communicative infrastructure, and drivers and motives, which are below discussed in turn.

3.1 Actors

Following a classic communication model, actors engaged in disinformation ecosystems are divided into four broad categories: *senders*, *receivers*, *mediators*, and *regulators*.

Senders of disinformation comprise a diversity of potential actors, both on an aggregate and individual scale. They include states, political parties, commercial interests, and their leaders, bots and employees. Among disinformation senders with a (geo)political strategic motive are 'cyber troops' that develop sometimes quite elaborate organisational forms, including strategies and capacities, to effectively manipulate public opinion in a desired direction (Bradshaw and Howard, 2017). Depending on the specific circumstances, cyber troops may engage private contractors, paid citizens, or volunteers (ibid: 15-6). Disinformation strategies are geared towards questioning established facts (i.e., commonly accepted observations of reality) or truths (intersubjectively accepted understandings of reality, grounded in facts). They do so principally by casting doubt on the trustworthiness of facts and/or truths, and replacing them by falsities purporting to be facts, and truths devoid of references to facts (Conrad, 2020: 9).

Ironically, in order to be successful at this endeavour, disinformation senders first require the audience to trust their facts and truths (at least more than those they criticise) (Michailidou et al., 2020: 6). They do so through a variety of affect-enhancing methods such as the invocation of nationalistic or opposition-favouring

imagery, attacks on government communication or other widely recognised institutions of public information, or harassment or threats to others expressing dissenting views (Bradshaw and Howard, 2017: 9). The positive messaging strategy (i.e., presentation of competing facts and truths) is complicated by the fact that potential receivers of their information have different views and preferences. This can be remedied by a focus on negative communication (i.e. discrediting established facts and truths). Disinformation senders exploit two factors in this regard. First, they can play up the (correct) observation that journalists and public authorities also do not have direct access to the truth, and represent 'partial', and sometimes poorly argued viewpoints as well (Michailidou et al., 2020: 3; Tambini, 2017: 3). With regard to traditional societal institutions of 'fact-finding', such as science and the judiciary, they can suggest partiality as well, by associating these institutions with government power, thus putting into doubt the reliability of their fact-finding activities (Dahlgren, 2018). Second, they can build on communication techniques designed to trigger emotions, particularly sensationalism, scandalisation, and ridicule, thence causing curiosity, outrage, or a sense of community amongst targeted audiences (Conrad, 2020; Vasu et al., 2018; Stark and Stegmann, 2020: 29).

Receivers of disinformation can in principle be any member of the public consuming information through media, either online or offline. However, since communications in offline media are often more strictly curated and disaggregated than those on online (particularly social) media, receivers are more likely to first encounter disinformation, which generally lacks the publicity of mainstream communication, via the latter channels (Stark and Stegmann, 2020: 34). On social media platforms, users often consume news shared by acquaintances, offering disinformation 'a way in' and an opportunity to spread. When information consumers are exposed more frequently to disinformation, they are more likely to give it consideration (Vasu et al., 2018: 16; Del Vicario et al., 2016: 554).

Information consumers may show different degrees of receptiveness to disinformation, depending on their attitude towards central institutions of democratic society, such as the government, large political parties, judiciary, media, and educational and research institutions (Dahlgren, 2018). Receivers' trust in these institutions' communication is based on trust in, and agreement with, the epistemology (knowledge base) or even ontology (notion of reality) that they propound (Michailidou et al., 2020: 3). A first step in undermining this trust is thus to reveal that these institutions (particularly so-called 'legacy media') only represent 'a truth', and that other truths also deserve a fair hearing (ibid: 4-5). Disinformation may convince receivers that the truths presented by democratic society's central institutions are flawed, because they are based on corrupted, self-serving, or invisible interests and motives. This increases the likelihood that they develop an attitude of rejection towards compelling counter-evidence to the disinformation narrative (Vasu et al., 2018: 16). In the typology of information attitudes presented by MacMullen (in Conrad, 2020), the consumption pattern of the disinformation receiver shifts from unconscious (deceived), to epistemic (hyper-relativistic), to motivational (unreceptive to the facts, potentially recruitable as a volunteer for the 'sender's side').

Sender and receiver however are only one side of the actor constellation in the disinformation ecosystem. An important further actor are *mediators*. In the case of online communication, senders are afforded the possibility of communicating directly (that is, without being curated) to an audience. However, contrary to what some scholars claim, this does not amount to full 'disintermediation', in the sense that "senders can reach receivers without go-between" (Del Vicario et al., 2016: 554). Instead, senders are still largely

dependent on mediators, such as social media platforms (most dominantly, Facebook, Twitter, and Instagram) or search engines (particularly Google). By managing the availability, accessibility, and prominence of communications, these mediators fulfil a gatekeeper function in the information ecosystem that gives them extraordinary influence over information communication patterns (MacGonagle, 2017: 206; Dahlgren, 2018: 21; Stark and Stegmann, 2020: 9). This influence is enhanced by the fact that many people consume their news via social media (Fletcher et al., 2018: 1; Dahlgren, 2018: 25), while as of January 2021, Google's search engine has a market share of over 86% (Statista, 2021).

Also media companies, both online and offline act as mediators of disinformation. A characteristic these mediators share is their for-profit operating model, which means that they largely earn their money from the advertisement revenue generated by audience attention (Stark and Stegmann, 2020: 9). In this sense, they favour communication styles and contents that enhances audiences, including controversial, scandalising, and antagonising information. This preference influences the selection of information selected or prioritised for audience visibility, in the case of online mediators, via algorithmic editorialisation, and for offline media, for selection and prominence of featured items. In the case of media companies, the frenetic news cycle and pressure to produce a constant stream of new content creates a further incentive to accept and publish low-cost content. This creates both a potential outlet for disinformation senders, and offers them a potential source of cases of poor reporting for the enhancement of their disinformation narrative.

Finally, there are *regulators* of communication channels. As public authorities have become more aware of the disinformation phenomenon due to a series of high-profile controversies in recent years (inter alia, 'Pizzagate', allegations of Russian, Chinese, and Indian interference in various national (electoral) public debates, and Trump's removal from Twitter), they have become more active in regulating online communication streams. An argument for this is that a free public debate on the basis of readily available factually correct information is essential to the proper functioning of democracy, and that guaranteeing the safe exercise of the freedom of speech is a primary task of state authority. For these reasons, proponents of active regulatory intervention argue that "[i]nformation defence cannot be outsourced to the public" (Aro, 2016: 130).

Public authorities have developed various strategies to tackle disinformation through regulation. The most important of these is to outsource regulation to social media platforms themselves, either on a voluntary basis (various social media platforms), or through regulatory coercion (e.g. Germany). Thus, Facebook has started to experiment with a feature in which controversial news reports are surrounded by 'related articles', while Twitter has recently started to add qualifying statements next to controversial tweets with a large reach. But public authorities can also take more proactive measures themselves, ranging from reporting portals for ensuring communicators' compliance with the law (Italy) to the identification of disinformation through a cross-country stakeholder network and regular publication of a disinformation digest (the EU). In Germany, social media platforms are bound by law to remove manifestly illegal material within 24 hours (Bode and Vraga, 2015; Alemanno, 2018: 3-4). Heavy-handed approaches involving repression have however met with criticism for their chilling effect on the freedom of expression and for putting public authorities in the undesirable position of policing the information sphere. Furthermore, this potentially also adds further conspiratorial grist to the disinformation mill.

3.2. Communicative infrastructure

In the account of Nardi and O'Day (1999), technology forms part of the information ecology, referring to the variety of tools, simple and advanced, at actors' disposal. Such tools do not only enable, but also shape information-related activities and understandings. In relation to disinformation creation and consumption, communication comes in many manifestations, including so-called 'troll farms' (private firms operating individuals and bots who engage on social media), news agencies (producing false news items for dissemination), newspapers, podcasts, vlogs, blogs, apps, and even official (government) webpages (Peters, 2018: 1162; Bradshaw and Howard, 2017: 11). In some cases, notably Russia and China, disinformation initiatives operate with clear (geo)political objectives, supported by considerable organisational capacity and financial resources (Bradshaw and Howard, 2017: 22). When targeting western democracies, such operations can be skilful at exploiting the relatively open media environment of these countries (Peters, 2018: 1162). This refers in the first place to online media.

Looking at the presence of online disinformation, the first thing that draws attention is its considerable fragmentation. A study of France and Italy, for example, indicates that even the most popular 'fake news' websites in 2017 reached no more than 3.5% of internet users, with by far most staying under 1% coverage. Moreover, on average, users spent far less time on such websites than on their mainstream counterparts. However, when disinformation connects to social media, its exposure increases considerably (Fletcher et al. 2018: 1; Stark and Stegmann, 2020: 45). A study of shares and likes of disinformation links on Facebook in France for example, indicates that while such links still have a comparatively limited reach compared to mainstream media links, a small number of disinformation outlets actually match or even exceed the mainstream counterparts (Fletcher et al. 2018: 4-5). Moreover, individual disinformation links can have a 'long tail', meaning that they linger on social media, reaching relatively large cumulative audiences over time (ibid: 7; also Del Vicario et al., 2016).

In terms of the specific novelty of online technology, particularly the scale, speed, and personalisation of disinformation are often mentioned (MacGonagle, 2017: 206; Stark and Stegmann, 2020). Scale and speed are most of all attributed to the 'portability' of disinformation, which derives both from the ease with which it can be 'copy-pasted' or 'shared', and from the platform function of social media that brings together wide audiences. Disinformation, aided by popularity-detecting social media algorithms, can attain 'virality', causing it to travel quickly through personal networks (Facebook 'friends' and Twitter 'followers') and networks of strangers (e.g. retweets on Twitter or Facebook groups) (Corner, 2017: 1102; Alemanno, 2018: 1). At the same time, virality can also be artificially created through 'cyber armies' consisting of humans and bots (Vasu et al., 2018: 12). Given the sheer amount of information passing through social media, users become accustomed to the fact that they cannot readily detect the origin or veracity of the information that reaches them, or the authenticity of the virality that caused it to reach them in the first place (Del Vicario et al., 2016: 554; MacGonagle, 2017: 206).

Besides fragmented and fast-moving, the technological profile of disinformation is also amorphous. Indeed, the extent of factual inaccuracy of a sender often depends on the site of communication and the target audience (Conrad, 2020: 9). Online, artificial intelligence, feeding partly on personal data sold by social media companies, develops predictive models individuals' views and preferences. This allows for tailor-made messaging in homogenous information environments, popularly known as 'filter bubbles' (Vasu et al., 2018:

10-1). When information consumers are convinced to distrust specific types of information or media, they increasingly retreat in such groups of like-minded persons. In 'filter bubbles', users screen out undesired or untrusted information, and share communications that confirm collectively held believes (Del Vicario et al., 2016: 554). These sites thus turn into 'echo chambers' that become alienated from the rest of the online information sphere (Dahlgren, 2018: 25). Incidentally, while this phenomenon is aided and accelerated by algorithmic content prioritisation, the homogeneous and antagonistic nature of filter bubbles in fact *decreases* the potential for virality (Del Vicario et al., 2016: 556; Stark and Stegmann, 2020: 14).

The specific technological environment in which disinformation operates appears to create certain psychological effects upon its receivers that enhance its effectiveness (Corner, 2017: 1103). Online information consumers, used to confronting large amounts of highly diverse and often irrelevant information in a single social media format (e.g. Facebook or Instagram timelines filled, next to news articles, with memes, private photos, video clips, jokes, and so on) are primed to take in information indifferently, expecting fast emotional triggers and easily discernible information cues (Corner, 2017: 1106). As a result, initial information intake is typically passive and trigger-prone. Furthermore, the relative anonymity and diversity of items also creates cognitive overflow. Because of this, the precise contents, time, and place of information initially observed are quickly forgotten. This may create an illusory truth effect, by which repeatedly presented communications are assessed to be credible in spite of their murky origins (Vasu et al., 2018: 15).

More generally, some evidence has been presented for the hypothesis that the overall acceleration of information streams in late modern culture structurally affects people's critical faculties (Dahlgren, 2018: 22-3). Vasu and colleagues (2018: 14) for example cite a 2009 meta-study by development psychologists of the UCLA which concludes that while growing internet use improves certain skills, it weakens "deep processing" capacities required for "mindful knowledge acquisition, inductive analysis, critical thinking, imagination, and reflection". Various authors therefore recommend online media literacy to be incorporated in educational curricula to decrease information consumers' vulnerability to these psychological trappings, although not all agree that this solution have a sufficient impact for the challenges that disinformation poses in the present (Vasu et al, 2018: 24; Martens et al., 2018: 49-50; Tambini, 2017: 15; Corner, 2017: 1102).

3.3 Drivers and motives

From the above discussion, it becomes clear that disinformation senders are active through a variety of online and offline information technologies to further their goal. Depending on the specific actor in question, this goal can be to mobilise ideological supporters and voters, to earn money, or to discredit specific actors and/or viewpoints (MacGonagle, 2017: 204-5; Conrad, 2020: 2; Vasu et al., 2018: 8). Disinformation senders employ various communicative styles and techniques to make these objectives attractive to news consumers, including the mobilisation of 'clickbait' (short messages that trigger people to click a link), 'astroturfing' (making posts appear to be grassroots), the use of bots, sponsored contents, satire, pseudoscience, propaganda, and account theft (MacGonagle, 2017: 204; Corner, 2017: 1103; Bradshaw and Howard, 2017: 10-2; Peters, 2018: 1162).

The disinformation method typically consists of three components: (1) attacking mainstream truths and facts based on (partially) made-up sources including incorrect facts, (2) presenting contorted counter-truths based on (partially) made-up sources including incorrect facts, and (3) silencing potential other senders who defend

mainstream truths or question the counter-truths. This sequence of activities underlines the malignancy of disinformation senders. In their endeavours, they rely on defamation and privacy breaches (including sensitive personal information and false suggestions concerning the psychological stability and moral character of prominent individuals) (Tucker et al., 2018: 24, 31; Bennett and Livingston, 2018: 125), deception (presenting false or falsely represented facts, either deliberately or by ignoring standards of due diligence) (Peters, 2018: 1162; MacGonagle, 2017: 208), and undermining the preconditions for free and open public debate (through harassment, threats, or obstruction, e.g. 'hashtag poisoning') (Aro, 2016; Bradshaw and Howard, 2017: 10-2). In the words of Peters (2018: 1132), "Information is the new warfare both against civil society and other countries". Indeed, increasingly, communications are not immediately damaged by demonstrations of their factual incorrectness (Conrad, 2020: 6).

The success of the disinformation formula can to an important extent be explained by a structural distrust in "democratic institutions of press and politics" among a growing minority of the population (Bennett and Livingston, 2018: 127; Michilaidou et al., 2020; Dahlgren, 2018; Peters, 2018: 1162). The rejection of mass media is correlated with age and educational level (both more likely when lower), and particularly associated with right-wing political preferences and contexts with strong political polarisation (Martens et al., 2017: 213). Some observers argue that the stage for the breakdown in trust was set "by decades of systematic deception on the part of power elites" (Dahlgren, 2018: 24). Others identify highly specific moments of breakdown of authority, such as the Iraq war and the 2008 financial crisis (D'Ancona in Corner, 2017: 1104). While the former representation lacks precision, the latter likely has limited explanatory value due to its local contingency. What seems to be nearly universally true in western democracies however, is a growing perception of mass media as actively 'constructing' the news, which may cause a more cynical attitude towards any invocation of 'the facts' and 'the truth' (Dahlgren, 2018: 22). At the same time, the suggestion of a 'post-truth era' incorrectly implies that there was a golden age of truth politics (Conrad, 2020: 5; Corner, 2017: 1100). Moreover, truth and fact as markers in communication have not disappeared: even senders of disinformation invoke them (Corner, 2017: 1106, see also Michailidou et al., 2020). Rather, a situation arises in which competing notions and institutions of knowledge and facts establish themselves, with little room for exchanges of arguments (Dahlgren, 2018: 22).

Polarisation in combination with the technological possibilities of social media leads to segregated information communities described by Tucker et al. (2018: 31) as "mutual admiration societies", or 'filter bubbles' as mentioned above. Such communities of shared views decrease the application of critical faculties to inward communication, leading to the uncritical acceptance of false information, hyper-partisan framing, and hostile and paranoid worldviews (Del Vicario et al., 2016: 558). Extreme distrust leads disinformation consumers to replace expert opinion by alternative, highly idiosyncratic readings of expert information, giving them a sense of personal authority and enhanced autonomy (Paul Levinson in Corner, 2017: 1102). Once a person has become invested in a strongly homogenous community of truth dissemination, he becomes "highly resistant to correction", as he is encouraged to engage with every argument countering the group's views in a distrustful or even hostile manner (Del Vicario et al., 2016: 558). As most energy is spent on protecting one's views and attack those of the other, the fact-orientation of disinformation audiences becomes blunted (Corner, 2017: 1101). Truth orientation is also 'tribalised'. Rather than exposing claims to rational scrutiny, the leading criterion of veracity becomes its "authenticity" (Dahlgren, 2018: 26). The

emotional appeal of communication thus increases at the expense of factual precision; more or less consciously, receivers want to 'feel good' about what they read (Corner, 2017: 1103).

The current predicament of the public information sphere is complex and, as indicated above, likely partially self-inflicted by the central institutions of democratic societies. Thus, while several authors suggest policy steps to counter the disinformation phenomenon (see particularly Tambini, 2017: 13-16; Haciyakupoglu et al., 2018: 8-12; Martens et al., 2017: 47-52; Bennett and Pfetsch, 2018: 247-50; Vasu et al., 2018: 18-25), many also caution against ill-conceived or repressive counter-measures. These critique comes in four varieties:

- First, it is held that strong government intervention can have *negative side-effects that outweigh the benefits*. For example, active state policing of disinformation can lead to 'Ministries of Truth' with unclear objectives and criteria of application, and strict enforcement can have a chilling effect on freedom of speech, which is the very things that the liberal democratic state purports to protect (Alemanno, 2018: 3). In this light, MacGonagle (2017: 205) rhetorically asks: "How much room should the watchdog be given to roam or how tight should the leash be?".
- Second, it is held that many proposals are likely to be *ineffective or even counterproductive*. For example, it is pointed out that readers of fake news are unlikely to look up fact-checking sites to verify the veracity of news stories, and that their reach will therefore remain limited (Vasu et al., 2018: 19). The 'fake news' label can furthermore give disinformation stories publicity and lend them virality. In any event, interventions against disinformation often come after they have spread widely (Alemanno, 2018: 4). More generally, it is suggested that an appeal to rationality alone is unlikely sway many sceptical of government and legacy media publicity (MacGonagle, 2017: 208).
- Third, it is argued that the *proposed measures leave an underlying problem untouched*. A particular problem mentioned in this regard is the business model of private social media companies, which economically incentivise the spread of misinformation (Alemanno, 2018: 5; Martens et al., 2018: 35 and further), as well as economic pressures on mass media, which lead to poor quality and factual errors and misrepresentations (Rabin-Havt in Corner, 2017: 1105). More generally, it is pointed out that discontented citizens are unlikely to give credence to sources that they do not know or actively mistrust. Misgivings are particularly deeply rooted where they stem from political disillusionment related to economic and socio-cultural policies (Kofman, 2018). Disgruntled citizens see little sense in trusting information that comes from those sources that they perceive to actively undermine their interests and values (Dahlgren, 2018; Conrad, 2020).
- Fourth, some studies argue that *the problem of disinformation is either overstated or inconclusive* and in need of further investigation. These studies for example point out the scant evidence of the wide spread or impact of specific disinformation narratives, the unlikeliness of their spreading beyond small homogenous groups, and the empirical indications that even most readers of disinformation consume at least as much, if not more mainstream news (Fletcher et al., 2018; Del Vicario et al., 2016; Michailidou et al., 2020; Stark and Stegmann, 2020: 35-6).

4. Conclusion

At the outset of this note, three leading questions were asked, to which we return here.

4.1 (How) does online communication alter the nature and functioning of disinformation?

While the literature indicates that efforts to defame and harass opponents and deceive publics for political and economic purposes are far from new, it does suggest that the jump of this phenomenon to the online sphere, and particularly to social media, has brought about certain significant changes. In particular, the scale and speed of circulation have increased, as has the potential for personalised contents due to new technological features. At the same time, it must be emphasised that also today's disinformation phenomena are multi-medial, employing both offline and online media. Studies of the psychological effects of social media engagement moreover indicate that certain characteristics of such engagement enhance users' receptiveness to well-crafted disinformation communications, and blunt users' critical faculties. Although the precise impact of disinformation remains contested, it seems clear that the phenomenon has enjoyed a constantly growing interest in the academic research.

4.2 How do the institutions for creating (and maintaining trust in) public information relate to disinformation?

The relation between disinformation and institutions for creating public information is predatory. Disinformation is more successful when trust in such institutions is low. At the same time, disinformation communications typically also incorporate distrust-fomenting messaging as part of their overall strategy. The central institutions of democratic society, including government, the largest media (including public and private broadcasters and the largest daily newspapers) and the individuals associated with them, as well as their representations of facts and truths are often the object of disinformation attacks. The alternative truth representations of disinformation are often created in juxtaposition to them. This suggests that much disinformation either serves interests that oppose the power or policies of these central institutions, or presupposes (an interest in) such opposition among target audiences. It furthermore appears that among audiences receptive to disinformation, distrust in the central institutions of democratic societies is fomented by longstanding antagonism towards various substantive policies and (perceived) failings of these institutions, which spills over into epistemological –sometimes even ontological– opposition. Disinformation thus often represents an explicitly partisan, *negative epistemology* that invests more scrutiny in opposing 'the official narrative' than in developing a detailed, coherent, and –for outsiders- convincing counter-narrative. In some cases, this leads to a breakdown in the possibility of democratic deliberation and debate. Particularly where disinformation senders are associated with hostile foreign powers, this may be the very intention. In the case of social media platforms with a profit incentive, it may be a perverse and unpremeditated side effect.

4.3 How do motives other than malignant intentions cause or exacerbate the disinformation phenomenon?

In cases of highly organised disinformation campaigns, intentions may be malignant and related to (geo)political objectives. However, disinformation thrives where there is a pre-existing disconnect between the central institutions of democratic society and a sizeable minority of the population, which causes its tailor-made messages to resonate. In particular, economic or socio-cultural alienation of parts of the community, as well as manifest failings (such as miscarriages of justice, political scandals, economic mismanagement, or gross journalistic omissions or misreporting) may have increased the receptiveness for disinformation, while providing the latter with plausible accusatory arguments. Beyond that, certain technological innovations associated with current-day online communication are also likely to have exacerbated the disinformation phenomenon. Among these are the easy of portability of online information,

the concentration of large communities of information consumers on social media such as Facebook, Instagram, and Twitter, and the functionalities on these social media, including sharing, 'liking', and response functions. However, possibly of greater influence still is the business model of these social media, which is based on advertisement revenue for consumer exposure. This model incentivises such social media companies to develop algorithms that increase users' 'screen time'. As a consequence, these algorithms prioritise emotion-triggering (scandalising, shocking, satirising and related) contents, which plays into the hand of polarising disinformation senders. Furthermore, many social media companies also sell their users' personal data, which allows third parties, including those with malignant intentions, to better craft and personalise their messages.

References

- Aro, J., 2016. The cyberspace war: propaganda and trolling as warfare tools. *European View* 15, 121-132.
- Allcott, H., and M. Gentzkow, 2017. Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives* 31(2), 211–236.
- Alemanno, A., 2018. Editorial: How to counter fake news? A taxonomy of anti-fake news approaches. *European Journal of Risk Regulation* 9(1), 1-5.
- BBC, 2021. Poland proposes social media 'free speech' law. https://www.bbc.com/news/technology-55678502. 15 January 2021.
- Bennett, W.L., and S. Livingston, 2018. The disinformation order: disruptive communication and the decline of democratic instutitions. *European Journal of Communication* 33(2), 122-139.
- Bennett, W.L., and B. Pfetsch, 2018. Rethinking Political Communication in a Time of Disrupted Public Spheres. *Journal of Communication* 68, 243-253.
- Bloomberg, 2021. Germany and France oppose Trump's Twitter exile. https://www.bloomberg.com/news/articles/2021-01-11/merkel-sees-closing-trump-s-social-media-accounts-problematic. 11 January 2021.
- Bode, L., and E.K. Vraga, 2015. In related news: that was wrong: the correction of misinformation through related stories functionality in social media. *Journal of Communication* 65, 619-638.
- Conrad, M., 2020. Post-truth politics as a mobilization tool: EU contestation in the Alternative for Germany's campaign for the 2019 European Parliament elections, paper prepared for the ECPR General Conference, 24-28 August 2020.
- Corner, J., 2017. Fake news, post-truth and media–political change. *Media, Culture & Society* 39(7), 1100-1107.
- Dahlgren, P., 2018. Media, knowledge and trust: the deepening epistemic crisis of democracy. *Javnost – The Public* 25(1-2), 20-27.
- Del Vicario, M. et al., 2016. The spreading of misinformation online. *PNAS* 113(3), 554-559.
- El-Barmawi, M.M., 2016. Your filter bubble is destroying democracy. *Wired Magazine*, 18 November 2016. https://www.wired.com/2016/11/filter-bubble-destroying-democracy/
- European Commission, 2017. Public consultation on fake news and online disinformation, 13 November 2017.
- European Commission, 2018. The digital transformation of news media and the rise of disinformation and fake news. JRC Digital Economy Working Papers 2018-02, April 2018.
- Fletcher, R., A. Cornia, L. Graves, and R. Klein Nielsen, 2018. Measuring the reach of 'fake news' and online disinformation in Europe. Reuters Institute Fact Sheet, February 2018.
- Guardian, 2018. New Facebook controls aim to regulate political ads and fight fake news. https://www.theguardian.com/technology/2018/apr/06/facebook-launches-controls-regulate-ads-publishers. 6 April 2018.
- Haciyakupoglu, G., J. Yang Hui, V.S. Suguna, D. Leong, and M.F. Bin Abdul Rahman, 2018. Countering fake news. A survey of recent global initiatives. S. Rajaratnam School of International Studies Policy Report. March 2018.
- Kofman, A., 2018. Bruno Latour, the post-truth philosopher, mounts a defense of science. *New Yorker*, 25 October 2018.
- Latour, B. 2005. *Reassembling the social: an introduction to actor-network theory.* Oxford: OUP.
- Martens, B., L. Aguilar, E. Gomez-Herrera, F. Mueller-Langer, 2018. The digital transformation of news media and the rise of disinformation and fake news. European Commission JRC Digital Economy Working Papers 2018-02.
- McGonagle, T., 2017. 'Fake news': False fears or real concern? *Netherlands Quarterly of Human Rights* 35(4), 203-209.
- Michailidou, A., H.J. Trenz, and E. Eike, 2020. Journalism and trust: the weaponisation of fake news in trust-building. *Forthcoming publication.*
- Nardi, B.A., and V.L. O'Day, 1999. Information Ecologies: Using Technology with Heart. MIT Press

- NBC, 2020. Twitter is testing new ways to fight misinformation – including a community-based points system. https://www.nbcnews.com/tech/tech-news/twitter-testing-new-ways-fight-misinformation-including-community-based-points-n1139931. 20 February 2020.
- Peters, M.A., 2018. The information wars, fake news, and the end of globalisation. *Educational Philosophy and Theory* 50(13), 1161-1164.
- Reuters, 2020. Facebook, Google, Twitter urged by EU to do more against fake news. https://www.reuters.com/article/eu-tech-disinformation/facebook-google-twitter-urged-by-eu-to-do-more-against-fake-news-idUKL8N2G6601. 10 September 2020.
- Stark, B., and D. Stegmann, 2020. Are algorithms a threat to democracy? The rise of intermediaries: a challenge for public discourse. Algorithms Watch Governing Platforms Report, May 2020.
- Statista, 2021. Worldwide desktop market share of leading search engines from January 2010 to January 2021. https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/
- Tambini, D., 2017. Fake news: public policy responses. LSE Media Policy Brief 20.
- Tucker, J.A., A. Guess, P. Barberá, C. Vaccari, A. Siegel, S. Sanovich, D. Stukal, and B. Nyhan, 2018. Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature. Report prepared for the Hewlett Foundation.
- Vasu, N., B. Ang, T.-A. Teo, S. Jayakumar, M. Faizal, and J. Ahuja, 2018. Fake news: national security in the post-truth era. S. Rajaratnam School of International Studies Policy Report. January 2018.